SFB 823

Optimal design for linear models with correlated observations

Holger Dette, Andrey Pepelyshev,
Anatoly Zhigljavsky

Discussion Paper

Nr. 39/2011

Submitted to the Annals of Statistics

OPTIMAL DESIGN FOR LINEAR MODELS WITH CORRELATED OBSERVATIONS

By Holger Dette*, Andrey Pepelyshev† and Anatoly Zhigljavsky‡

Ruhr-Universität Bochum, RWTH Aachen† and Cardiff University‡*

In the common linear regression model the problem of determining optimal designs for least squares estimation is considered in the case where the observations are correlated. A necessary condition for the optimality of a given design is provided, which extends the classical equivalence theory for optimal designs in models with uncorrelated errors to the case of dependent data. For one parameter models this condition is also shown to be sufficient in many cases and for several models optimal designs can be identified explicitly. For the multi-parameter regression models a simple relation which allows verifying the necessary optimality condition is established. Moreover, it is proved that the arcsine distribution is universally optimal for the polynomial regression model with a correlation structure defined by the logarithmic potential. It is also shown that for models in which the regression functions are eigenfunctions of an integral operator induced by the correlation kernel of the error process, designs satisfying the necessary conditions of optimality can be found explicitly. To the best knowledge of the authors these findings provide the first explicit results on optimal designs for regression models with correlated observations, which are not restricted to the location scale model.

1. Introduction. Consider the common linear regression model

$$y(x) = \theta_1 f_1(x) + \ldots + \theta_m f_m(x) + \varepsilon(x) \ , \tag{1.1}$$

where $f_1(x), \ldots, f_m(x)$ are given linearly independent functions, $\varepsilon(x)$ denotes a random error process or field, $\theta_1, \ldots, \theta_m$ are unknown parameters and x is the explanatory variable, which varies in a compact design space $\mathcal{X} \subset \mathbb{R}^d$. We assume that N observations, say $y_1, \ldots, y_N$, can be taken at experimental conditions $x_1, \ldots, x_N$ to estimate the parameters in the linear regression model (1.1). If an appropriate estimate, say $\hat{\theta}$, of the parameter $\theta = (\theta_1, \ldots, \theta_m)^T$ has been chosen, the quality of the statistical analysis can be further improved by choosing an appropriate design for the experiment. In particular, an optimal design minimizes a functional of the

AMS 2000 subject classifications: Primary 62K05, 60G10; secondary 31A10, 45C05

Keywords and phrases: optimal design, correlated observations, integral operator, eigenfunctions, arcsine distribution, logarithmic potential

variance-covariance matrix of the estimate $\hat{\theta}$, where the functional should reflect certain aspects of the goal of the experiment. In contrast to the case of uncorrelated errors, where numerous results and a rather complete theory are available [see for example the monograph of Pukelsheim (2006)], the construction of optimal designs for dependent observations is intrinsically more difficult. However, this problem is of particular practical interest as in most applications there exists correlation between different observations. Typical examples include models, where the explanatory variable x represents the time and all observations correspond to one subject. In such situations optimal experimental designs are very difficult to find even in simple cases. Some exact optimal design problems were considered in Boltze and Näther (1982), Näther (1985a), Ch. 4, Näther (1985b), Pázman and Müller (2001) and Müller and Pázman (2003), who derived optimal designs for the location scale model

$$y(x) = \theta + \varepsilon(x). \tag{1.2}$$

Exact optimal designs for specific linear models have been investigated in Dette et al. (2008a); Kiselak and Stehlik (2008); Harman and Štulajter (2010). Because explicit solutions of optimal design problems for correlated observations are rarely available several authors have proposed to determine optimal designs based on asymptotic arguments [see for example Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979), Näther (1985a), Zhigljavsky et al. (2010)]. Roughly speaking, there exist three approaches to embed the optimal design problem for regression models with correlated observations in an asymptotic optimal design problem. The first one is due to Sacks and Ylvisaker (1966, 1968), who assumed that the covariance structure of the error process $\varepsilon(x)$ is fixed and that the number of design points tends to infinity. Alternatively, Bickel and Herzberg (1979) and Bickel et al. (1981) considered a different model, where the correlation function depends on the sample size. Recently, Zhigljavsky et al. (2010) extended the Bickel-Herzberg approach and allowed the variance (in addition to the correlation function) to vary as the number of observations changes. As a result, the corresponding optimality criteria contain a kernel with a singularity at zero. The focus in all of these papers is again mainly on the location scale model (1.2).

The difficulties in the development of the optimal design theory for correlated observations can be explained by a different structure of the covariance of the least squares estimator in model (1.1), which is of the form $M^{-1}BM^{-1}$ for certain matrices M and B depending on the design. As a consequence the corresponding design problems are in general not convex (except for the location scale model (1.2) where $M = 1$).

The present paper is devoted to the problem of determining optimal designs for more general models with correlated observations than the simple location scale model (1.2). In Section 2 we present some preliminary discussion and introduce some notation. Section 3 is devoted to necessary conditions for design optimality. In Section 4 we prove the optimality of the arcsine distribution and a class of Beta distributions (called generalized arcsine designs) for specific one-parameter regression models. In Section 5 we derive necessary conditions for the universal optimality of designs in multi-parameter regression models with correlated errors. By relating the optimal design problems to eigenvalue problems for integral operators we identify a broad class of models where these conditions are satisfied. One of the main results of the paper is Theorem 5.2, where we prove that the arcsine design is universally optimal for the polynomial regression model with the logarithmic correlation kernel. To our best knowledge these results provide the first explicit solution of optimal design problems for regression models with correlated observations which differ from the location scale model. In Section 6 we provide an algorithm for computing optimal designs for any regression model with specified covariance function and investigate the efficiency of the arcsine and uniform distribution in polynomial regression models with exponential correlation functions. Finally, some conclusions are given in Section 7.

2. Preliminaries. Consider the linear regression model (1.1), where $\varepsilon(x)$ is a stationary process with

$$\mathrm{E}\varepsilon(x) = 0, \quad \mathrm{E}\varepsilon(x)\varepsilon(x') = \sigma^2 K(x, x'), \quad x \in \mathcal{X} \subset \mathbb{R}^d. \tag{2.1}$$

Throughout this paper we assume that the function $K(x, x')$ is continuous at all points $(x, x') \in \mathcal{X} \times \mathcal{X}$ except possibly the diagonal points (x, x). We also assume that $K(x, x') \neq 0$ for at least one pair (x, x') with $x \neq x'$. An important case appears when the correlation kernel of the error process is of the form $K(x, x') = \rho(x - x')$, where $\rho(\cdot)$ is called the correlation function. If N observations, say $y = (y_1, \ldots, y_N)^T$ are available at experimental conditions $x_1, \ldots, x_N$ and the knowledge of the correlation kernel is available, the vector of parameters can be estimated by the weighted least squares method, i.e. $\hat{\theta} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} y$ where $\mathbf{X} = (f_i(x_j))_{j=1,\ldots,N}^{i=1,\ldots,m}$ and $\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\ldots,N}$. The variance-covariance matrix of this estimate is given by

$$\mathrm{Var}(\hat{\theta}) = \sigma^2 (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1}.$$

If the correlation structure of the process is not known, one usually uses the ordinary least squares estimate $\tilde{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$, which has the covariance matrix

$$\operatorname{Var}(\tilde{\theta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}. \tag{2.2}$$

An exact experimental design $\xi_N = \{x_1, \dots, x_N\}$ is a collection of N points in $\mathcal{X}$, which defines the time points or experimental conditions where observations are taken. Optimal designs for weighted or ordinary least squares estimation minimize a functional of the covariance matrix of the weighted or ordinary least squares estimate, respectively, and numerous optimality criteria have been proposed in the literature to discriminate between competing designs [see Pukelsheim (2006)].

Note that the weighted least squares estimate can only be used if the correlation structure of errors is known, and its misspecification can lead to a severe loss of efficiency. On the other hand, the ordinary least squares estimate does not employ the structure of the correlation. Obviously the ordinary least squares estimate can be less efficient than the weighted least squares estimate but in many cases the loss of efficiency is often negligible. Throughout this article we will concentrate on optimal designs for the ordinary least squares estimate. These designs require also the specification of the correlation structure but a potential loss by its misspecification in the stage of design construction is typically much smaller than the loss caused by the misspecification of the correlation structure in the weighted least squares estimate [see Tables 1 and 3 in Dette et al. (2009)].

Because even in simple models the exact optimal designs are difficult to find, most authors usually use asymptotic arguments to determine efficient designs for the estimation of the model parameters. Following Sacks and Ylvisaker (1966, 1968) and Näther (1985a), Chapter 4, we assume that the design points $\{x_1, \dots, x_N\}$ are generated by the quantiles of a distribution function, that is

$$x_{iN} = a\Big((i-1)/(N-1)\Big),\ i = 1, \dots, N,$$

where the function $a : [0,1] \to \mathcal{X}$ is the inverse of a distribution function. If ξ_N denotes a design with N points and corresponding quantile function $a(\cdot)$, the covariance matrix of the least squares estimate $\tilde{\theta} = \tilde{\theta}_{\xi_N}$ given in (2.2) can be written as

$$\operatorname{Var}(\tilde{\theta}) = \sigma^2 D(\xi_N) = \sigma^2 M^{-1}(\xi_N)B(\xi_N, \xi_N)M^{-1}(\xi_N), \tag{2.3}$$

where

$$M(\xi_N) = \int_{\mathcal{X}} f(u) f^T(u) \xi_N(\mathrm{d}u), \tag{2.4}$$

$$B(\xi_N, \xi_N) = \iint K(u,v) f(u) f^T(v) \xi_N(\mathrm{d}u) \xi_N(\mathrm{d}v), \tag{2.5}$$

and $f(u) = \big(f_1(u), \dots, f_m(u)\big)^T$ denotes the vector of regression functions. Following Kiefer (1974) we call any probability measure ξ on $\mathcal{X}$ (more precisely on an appropriate Borel field) an approximate design or simply design. The set of all designs (that is, the set of all probability measures on $\mathcal{X}$) will be denoted by Ξ. The definition of the matrices $M(\xi)$ and $B(\xi, \xi)$ can be extended to an arbitrary design ξ, provided that the corresponding integrals exist. The matrix

$$D(\xi) = M^{-1}(\xi) B(\xi, \xi) M^{-1}(\xi), \tag{2.6}$$

is called the covariance matrix for the design ξ and can be defined for any probability measure ξ supported on the design space $\mathcal{X}$ such that the matrices $B(\xi, \xi)$ and $M^{-1}(\xi)$ are well-defined. An (approximate) optimal design minimizes a functional of the covariance matrix $D(\xi)$ over the set Ξ.

Note that in general the function $D(\xi)$ is not convex (with respect to the Loewner ordering) on the space of all approximate designs. This implies that even if one determines optimal designs by minimizing a convex functional, say Φ, of the matrix $D(\xi)$, the corresponding functional $\xi \to \Phi(D(\xi))$ is generally not convex on the space of designs Ξ. Consider for example the case $m = 1$ where $D(\xi)$ is given by

$$D(\xi) = \left[\int f^2(u) \xi(\mathrm{d}u) \right]^{-2} \iint K(u,v) f(u) f(v) \xi(\mathrm{d}u) \xi(\mathrm{d}v)\,, \tag{2.7}$$

and it is obvious that this functional is not necessarily convex. On the other hand, for the location scale model (1.2) we have $m = 1$, $f(x) = 1$ for all x and this expression reduces to $D(\xi) = \iint K(u,v) \xi(\mathrm{d}u) \xi(\mathrm{d}v)$. In the case $K(u,v) = \rho(u - v)$, where $\rho(\cdot)$ is a correlation function, this functional is convex on the set of probability measures on the domain $\mathcal{X}$, see Lemma 1 in Zhigljavsky et al. (2010) and Lemma 4.3 in Näther (1985a). For this reason (namely the convexity of the functional $D(\xi)$) most of the literature discussing asymptotic optimal design problems for least squares estimation in the presence of correlated observations considers the location scale model, which corresponds to the estimation of the mean of a stationary process [see for example Boltze and Näther (1982), Näther (1985a,b)].

3. A necessary condition for optimality.

3.1. *General optimality criteria.* Recall the definition of the information matrix in (2.4) and define

$$B(\xi,\nu)=\int_{\mathcal{X}}\int_{\mathcal{X}}K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\nu(\mathrm{d}v),$$

where ξ and $\nu\in\Xi$ are two arbitrary designs and $K(u,v)$ is an arbitrary kernel. The two main examples of the kernel function K will be $K(u,v)=\rho(u-v)$ and $K(u,v)=r(u-v)$ where $\rho(t)$ is a correlation function and $r(t)$ is a non-negative definite function with singularity at zero. The latter type arises naturally if the Bickel-Herzberg approach [see Bickel and Herzberg (1979)] is extended such that the variance (in addition to the correlation function) varies as the number of observations changes [see Zhigljavsky et al. (2010)] .

According to the discussion in the previous paragraph, the asymptotic covariance matrix of the least squares estimator $\hat{\theta}$ is proportional to the matrix $D(\xi)$ defined in (2.3). Let $\Phi(\cdot)$ be a monotone functional defined on the space of symmetric $m\times m$ matrices where the monotonicity of $\Phi(\cdot)$ means that $A\geq B$ implies $\Phi(A)\geq\Phi(B)$. Then the optimal design ξ^* minimizes the function

$$\Phi(D(\xi)) \tag{3.1}$$

on the space Ξ of all approximate designs. In addition to monotonicity, we shall also assume the differentiability of the functional $\Phi(\cdot)$; that is, the existence of the matrix of derivatives

$$C=\frac{\partial\Phi(D)}{\partial D}=\left(\frac{\partial\Phi(D)}{\partial D_{ij}}\right)_{i,j=1,\ldots,m},$$

where D is any symmetric non-negative definite matrix of size $m\times m$. The following lemma is crucial in the proof of the optimality theorem below.

Lemma 3.1 *Let ξ and ν be two designs and Φ be a differentiable functional. Set $\xi_\alpha=(1-\alpha)\xi+\alpha\nu$ and assume that the matrices $M(\xi)$ and $B(\xi,\xi)$ are nonsingular. Then the directional derivative of Φ at the design ξ in the direction of $\nu-\xi$ is given by*

$$\left.\frac{\partial\Phi(D(\xi_\alpha))}{\partial\alpha}\right|_{\alpha=0}=2[\mathbf{b}(\nu,\xi)-\boldsymbol{\varphi}(\nu,\xi)]$$

where

$$\boldsymbol{\varphi}(\nu,\xi)=\mathrm{tr}(M(\nu)M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi)M^{-1}(\xi)),$$

$$\mathbf{b}(\nu,\xi) = \mathrm{tr}(M^{-1}(\xi)\, C(\xi) M^{-1}(\xi) B(\xi,\nu))$$

and

$$C(\xi) = \left.\frac{\partial \Phi(D)}{\partial D}\right|_{D=D(\xi)}.$$

Proof. Straightforward calculation shows that

$$\left.\frac{\partial}{\partial \alpha} M^{-1}(\xi_\alpha)\right|_{\alpha=0} = M^{-1}(\xi) - M^{-1}(\xi) M(\nu) M^{-1}(\xi)$$

and

$$\left.\frac{\partial}{\partial \alpha} B(\xi_\alpha,\xi_\alpha)\right|_{\alpha=0} = B(\xi,\nu) + B(\nu,\xi) - 2B(\xi,\xi).$$

Using the formula for the derivative of a product and the two formulas above, we obtain

$$\begin{aligned}
\left.\frac{\partial}{\partial \alpha} D(\xi_\alpha)\right|_{\alpha=0} &= \left.\left(\frac{\partial}{\partial \alpha} M^{-1}(\xi_\alpha)\right) B(\xi_\alpha,\xi_\alpha) M^{-1}(\xi_\alpha)\right|_{\alpha=0} \\
&\quad \left. + M^{-1}(\xi_\alpha) B(\xi_\alpha,\xi_\alpha) \left(\frac{\partial}{\partial \alpha} M^{-1}(\xi_\alpha)\right)\right|_{\alpha=0} \\
&\quad \left. + M^{-1}(\xi_\alpha) \left(\frac{\partial}{\partial \alpha} B(\xi_\alpha,\xi_\alpha)\right) M^{-1}(\xi_\alpha)\right|_{\alpha=0} \\
&= -M^{-1}(\xi) M(\nu) M^{-1}(\xi) B(\xi,\xi) M^{-1}(\xi) \\
&\quad - M^{-1}(\xi) B(\xi,\xi) M^{-1}(\xi) M(\nu) M^{-1}(\xi) \\
&\quad + M^{-1}(\xi) \Big(B(\xi,\nu) + B(\nu,\xi)\Big) M^{-1}(\xi).
\end{aligned}$$

Note that the matrices $M(\xi_\alpha)$ and $B(\xi_\alpha,\xi_\alpha)$ are nonsingular for small non-negative α (that is, for all $\alpha \in [0,\alpha_0)$ where α_0 is a small positive number) which follows from the non-degeneracy of $M(\xi)$ and $B(\xi,\xi)$ and the continuity of $M(\xi_\alpha)$ and $B(\xi_\alpha,\xi_\alpha)$ with respect to α.

Using the above formula and the fact that $\mathrm{tr}(H(A+A^T)) = 2\,\mathrm{tr}(HA)$ for any $m\times m$ matrix A and any $m\times m$ symmetric matrix H, we obtain

$$\begin{aligned}
\left.\frac{\partial \Phi(D(\xi_\alpha))}{\partial \alpha}\right|_{\alpha=0} &= \left.\mathrm{tr}\Big(C(\xi)\frac{\partial}{\partial \alpha} D(\xi_\alpha)\Big)\right|_{\alpha=0} \\
&= 2\,\mathrm{tr}\Big(C(\xi) M^{-1}(\xi) B(\xi,\nu) M^{-1}(\xi)\Big) \\
&\quad -2\,\mathrm{tr}\Big(C(\xi) M^{-1}(\xi) M(\nu) M^{-1}(\xi) B(\xi,\xi) M^{-1}(\xi)\Big) \\
&= 2[\mathbf{b}(\nu,\xi) - \boldsymbol{\varphi}(\nu,\xi)]\,.
\end{aligned}$$

□

Note that the functions $\mathbf{b}(\nu,\xi)$ and $\boldsymbol{\varphi}(\nu,\xi)$ can be represented as

$$\mathbf{b}(\nu,\xi) = \int b(x,\xi)\nu(\mathrm{d}x), \quad \boldsymbol{\varphi}(\nu,\xi) = \int \varphi(x,\xi)\nu(\mathrm{d}x)$$

where

$$\begin{aligned} (3.2)\ \varphi(x,\xi) &= \boldsymbol{\varphi}(\xi_x,\xi) = f^T(x)M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi)M^{-1}(\xi)f(x) \\ (3.3)\ \ b(x,\xi) &= \mathbf{b}(\xi_x,\xi) = \mathrm{tr}(C(\xi)M^{-1}(\xi)B(\xi,\xi_x)M^{-1}(\xi)) \end{aligned}$$

and ξ_x is the probability measure concentrated at a point x.

Lemma 3.2 *For any design ξ such that the matrices $M(\xi)$ and $B(\xi,\xi)$ are nonsingular we have*

$$(3.4) \int \varphi(x,\xi)\xi(\mathrm{d}x) = \int b(x,\xi)\xi(\mathrm{d}x) = \mathrm{tr}(M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi))$$

where the functions $\varphi(x,\xi)$ and $b(x,\xi)$ are defined in (3.2) *and* (3.3)*, respectively.*

Proof. Straightforward calculation shows that

$$\begin{aligned} \int \varphi(x,\xi)\xi(\mathrm{d}x) &= \mathrm{tr}(M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi)M^{-1}(\xi)\int f(x)f^T(x)\xi(\mathrm{d}x)) \\ &= \mathrm{tr}(M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi)). \end{aligned}$$

We also have

$$\begin{aligned} \int B(\xi,\xi_x)\xi(\mathrm{d}x) &= \int \Big[\int\int K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\xi_x(\mathrm{d}v)\Big]\xi(\mathrm{d}x) \\ &= \int \Big[\int K(u,x)f(u)f^T(x)\xi(\mathrm{d}u)\Big]\xi(\mathrm{d}x) = B(\xi,\xi), \end{aligned}$$

which implies

$$\begin{aligned} \int b(x,\xi)\xi(\mathrm{d}x) &= \mathrm{tr}(M^{-1}(\xi)\,C(\xi)M^{-1}(\xi)\int B(\xi,\xi_x)\xi(\mathrm{d}x)) \\ &= \mathrm{tr}(M^{-1}(\xi)\,C(\xi)M^{-1}(\xi)B(\xi,\xi)) \\ &= \mathrm{tr}(M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi)). \end{aligned}$$

□

The main result of this section provides a necessary condition for the optimality of a given design.

Theorem 3.1 *Let ξ^* be any design minimizing the functional $\Phi(D(\xi))$. Then the inequality*

$$\varphi(x, \xi^*) \leq b(x, \xi^*) \tag{3.5}$$

holds for all $x \in \mathcal{X}$, where the functions $\varphi(x, \xi)$ and $b(x, \xi)$ are defined in (3.2) *and* (3.3)*, respectively. Moreover, there is equality in* (3.5) *for ξ^*-almost all x, that is, $\xi^*(\mathcal{A}) = 0$ where*

$$\mathcal{A} = \mathcal{A}(\xi^*) = \{x \in \mathcal{X} \mid \varphi(x, \xi^*) < b(x, \xi^*)\}$$

is the set of $x \in \mathcal{X}$ such that the inequality (3.5) *is strict.*

Proof. Consider any design ξ^* minimizing the functional $\Phi(D(\xi))$. The necessary condition for an element to be a minimizer of a differentiable functional states that the directional derivative from this element in any direction is non-negative. In the case of the design ξ^* and the functional $\Phi(D(\xi))$ this yields that for any design ν

$$\left. \frac{\partial \Phi(D(\xi_\alpha))}{\partial \alpha} \right|_{\alpha=0} \geq 0$$

where $\xi_\alpha = (1 - \alpha)\xi^* + \alpha\nu$. The inequality (3.5) follows now from Lemma 1. The assumption that the inequality (3.5) is strict for all $x \in \mathcal{A}$ with $\xi^*(\mathcal{A}) > 0$ is in contradiction with the identity (3.4). □

Remark 3.1 In the classical theory of optimal design, convex optimality criteria are almost always considered. However, in at least one paper, namely Torsney (1986), an optimality theorem for a rather general non-convex optimality criteria was established and used (in the case of non-correlated observations).

3.2. *Special optimality criteria.* For the L-optimality criterion defined by $\Phi(D) = \text{tr}(LD)$, where L is a symmetric non-negative matrix, we have $C(\xi) = L$ for any ξ. The formulas above do not simplify much. In a particular case of the A-optimality criterion $\Phi(D) = \text{tr}\, D$ we have $C(\xi) = I$ for any ξ. For the case of c-optimality criterion $\Phi(D) = c^T Dc = \text{tr}(cc^T D)$ we obtain $C(\xi) = cc^T$ for any ξ. In both cases, simplifications in the expressions for $\varphi(x, \xi)$ and $b(x, \xi)$ are insignificant.

For the D-optimality there exists an analogue of the celebrated 'Equivalence Theorem' of Kiefer and Wolfowitz (1960), which characterizes optimal designs minimizing the D-optimality criterion $\Phi(D(\xi)) = \ln \det(D(\xi))$.

Theorem 3.2 *Let ξ^* be any D-optimal design. Then for all $x \in \mathcal{X}$ we have*

$$d(x, \xi^*) \le b(x, \xi^*), \tag{3.6}$$

where the functions d and b are defined by $d(x,\xi) = f^T(x)M^{-1}(\xi)f(x)$

$$b(x,\xi) = \mathrm{tr}\Big(B^{-1}(\xi,\xi)B(\xi,\xi_x)\Big) = f^T(x)B^{-1}(\xi,\xi)\int K(u,x)f(u)\xi(\mathrm{d}u), \tag{3.7}$$

respectively. Moreover, there is equality in (3.6) *for ξ^*-almost all x.*

Proof. In the case of the D-optimality criterion $\Phi(D(\xi)) = \ln\det(D(\xi))$, we have

$$C(\xi) = D^{-1}(\xi) = (M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi))^{-1} = M(\xi)B^{-1}(\xi,\xi)M(\xi), \tag{3.8}$$

which gives

$$\varphi(x,\xi) = f^T(x)M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)M(\xi)B^{-1}(\xi,\xi)M(\xi)M^{-1}(\xi)f(x) = d(x,\xi).$$

Similarly we simplify an expression for $b(x,\xi)$. Reference to Theorem 1 completes the proof. □

The following statement illustrates a remarkable similarity between D-optimal design problems in the cases of correlated and non-correlated observations. The proof easily follows from the formulas (3.4) and (3.8).

Corollary 3.1 *For any design ξ such that the matrices $M(\xi)$ and $B(\xi,\xi)$ are nonsingular we have*

$$\int d(x,\xi)\xi(\mathrm{d}x) = \int b(x,\xi)\xi(\mathrm{d}x) = m$$

where $b(x,\xi)$ is defined in (3.7) *and m is the number of parameters in the regression model* (1.1).

As an illustration we consider the quadratic regression model $y(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \varepsilon(x)$ with design space $\mathcal{X} = [-1,1]$ and correlated observations. In Figure 1 we plot functions $b(x,\xi)$ and $d(x,\xi)$ for different covariance kernels $K(u,v) = e^{-|u-v|}$, $K(u,v) = \max\{0, 1-|u-v|\}$ and $K(u,v) = -\log(u-v)^2$, where the design is the arcsine distribution with density

$$p(x) = 1/(\pi\sqrt{1-x^2}), \quad x \in (-1,1). \tag{3.9}$$

Throughout this paper this design will be denoted by ξ_a. By the definition, the function $d(x,\xi)$ is the same for different covariance kernels but the function $b(x,\xi)$ depends on the choice of the kernel. From the left and middle panel we see that the arcsine distribution does not satisfy the necessary condition of Theorem 3.1 for the kernel $K(u,v) = e^{-|u-v|}$ and $\max\{0, 1-|u-v|\}$ and is therefore not D-optimal for the quadratic regression model. On the other hand, for the logarithmic kernel $K(u,v) = -\log(u-v)^2$ the necessary condition is satisfied and the arcsine distribution is a candidate for the D-optimal design. We will prove in Theorem 5.2 that the arcsine design ξ_a is optimal with respect to a broad class of optimality criteria including the D-optimality criterion.

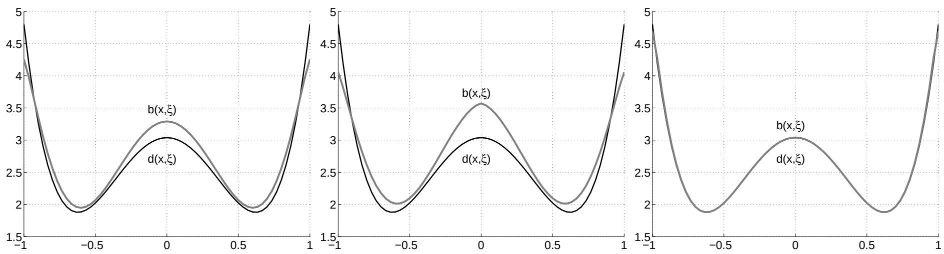

FIG 1. *The functions $b(x,\xi)$ and $d(x,\xi)$ for the regression model* (1.1) *with* $f(x) = (1, x, x^2)^T$ *and the covariance kernels* $K(u,v) = e^{-|u-v|}$ *(left),* $K(u,v) = \max(0, 1-|u-v|)$ *(middle) and* $K(u,v) = -\log(u-v)^2$ *(right), and the arcsine distribution* ξ_a.

4. Optimal designs for one-parameter models. In this section we study one-parameter regression models

$$y(x) = \theta f(x) + \varepsilon(x), \quad x \in \mathcal{X} \tag{4.1}$$

with correlated observations. It turns out that in this case a more explicit characterization of the optimal designs is available. Moreover, the results indicate a general strategy to deal with optimal designs in the general regression model (1.1) with correlated observations, which will be further developed in Section 5.

4.1. *Set of admissible designs.* For a one-parameter model of the form (4.1) $D(\xi)$ is a scalar and no functional Φ is needed to define an optimal design problem. In this case the optimality criterion reduces to

$$D(\xi) = \frac{B(\xi,\xi)}{(M(\xi))^2} = \left[\int f^2(u)\xi(\mathrm{d}u)\right]^{-2} \int\int K(u,v) f(u) f(v) \xi(\mathrm{d}u)\xi(\mathrm{d}v) \,. \tag{4.2}$$

This criterion has to be minimized on the set Ξ is of all admissible designs, which will be defined below.

Throughout this section we assume that the function $f(\cdot)$ is not identically zero so that the estimation of the parameter θ is possible. We also assume that the kernel $K(u,v)$ guarantees the existence of at least one design with $D(\xi) < \infty$ so that the problem of minimizing the functional $D(\xi)$ is well defined. For the following discussion define $\mathcal{X}_0 = \{x \in \mathcal{X} : f(x) = 0\}$ and $\mathcal{X}_1 = \mathcal{X} \backslash \mathcal{X}_0 = \{x \in \mathcal{X} : f(x) \neq 0\}$.

In the one-parameter case the (ordinary) least squares estimator $\tilde{\theta}$ based on observations $y_1, \ldots, y_N$ at points $x_1, \ldots, x_N$ is given by

$$\tilde{\theta} = (X^T X)^{-1} X^T Y = \frac{\sum_{i=1}^N f(x_i) y_i}{\sum_{i=1}^N f^2(x_i)}. \tag{4.3}$$

This estimator makes sense only if at least one of the predictors x_i belongs to the set $\mathcal{X}_1$. Otherwise, if all $x_i \in \mathcal{X}_0$, the parameter θ in (4.1) is not identifiable. Hence, in the formulation of the optimal design problem, all designs, that are concentrated entirely on the set $\mathcal{X}_0$, should be excluded from the set of admissible designs Ξ.

Assume now that at least one of the design points x_i belongs to the set $\mathcal{X}_1$. Then the variance of the estimate $\tilde{\theta}$ is given by

$$\operatorname{Var}(\tilde{\theta}) = \sigma^2 (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} = \sigma^2 \frac{\sum_{i,j} f(x_i) K(x_i, x_j) f(x_j)}{\left(\sum_i f^2(x_i)\right)^2}. \tag{4.4}$$

Any admissible design ξ can be represented as

$$\xi = \alpha \xi_1 + (1 - \alpha) \xi_0 \tag{4.5}$$

where $0 < \alpha \leq 1$ and ξ_k is the restriction of the measure ξ to the set $\mathcal{X}_k$, that is $\xi_k(A) = \xi(A \cap \mathcal{X}_k)/\xi(\mathcal{X}_k)$ for $k = 0, 1$ and any measurable set $A \subset \mathcal{X}$. Note that the case $\alpha = 0$ means that the design ξ is concentrated on the set $\mathcal{X}_0$; the corresponding designs have been excluded from the set of admissible designs.

The formulas (4.3) and (4.4) now imply that the observations $x_i \in \mathcal{X}_0$ do not change both the estimator $\tilde{\theta}$ and its variance $\operatorname{Var}(\tilde{\theta})$. That is, $D(\xi) = D(\xi_1)$ whatever the value of $\alpha \in (0, 1]$ is. This can also be checked directly using the formula (4.2); here we have to note that even if the kernel $K(u,v)$ has a singularity at $u = v$, we have $\iint K(u,v) f(u) f(v) \xi_0(\mathrm{d}u) \xi_0(\mathrm{d}v) = 0$. Since $D(\xi_1) = D(\alpha \xi_1 + (1 - \alpha) \xi_0)$ for all $\alpha \in (0, 1]$, we can assume that $\alpha = 1$.

Summarizing we assume that the set of admissible designs Ξ in the formulation of the optimal design problem (4.2) is a subset of all approximate designs defined on $\mathcal{X}$ containing all designs ξ such that $f(x) \neq 0$ for ξ-almost all $x \in \mathcal{X}$.

4.2. *Optimality conditions.* In the case of a one-parameter model, the functions $d(x,\xi)$ and $b(x,\xi)$ defined in (3.2) and (3.3) are given by

$$d(x,\xi) = \frac{f^2(x)}{\int f^2(u)\xi(\mathrm{d}u)}$$

$$b(x,\xi) = \frac{f(x)\int K(u,x)f(u)\xi(\mathrm{d}u)}{\iint K(u,v)f(u)f(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)},$$

respectively. In addition to the general necessary condition of optimality formulated in Theorem 3.1, we can formulate other conditions.

Theorem 4.1 *Consider a one-parameter model of the form* (4.1) *and an admissible design* $\xi \in \Xi$.

(i) If the design ξ *is optimal, then there exists* $\lambda > 0$ *such that the identity*

$$\lambda f(x) = \int K(u,x)f(u)\xi(\mathrm{d}u) \tag{4.6}$$

holds for ξ*-almost all* $x \in \mathcal{X}$.

(ii) If there exists $\lambda > 0$ *such that the condition* (4.6) *holds for all* $x \in \mathcal{X}$, *then the design* ξ *satisfies the necessary condition of optimality formulated in Theorem* (3.1).

(iii) If there exists exactly one pair (ξ,λ), *where* ξ *is an admissible design and* λ *is a positive scalar, such that* (4.6) *holds for* ξ*-almost all* $x \in \mathcal{X}$, *then* ξ *is the unique design minimizing the criterion* (4.2) *in the set of all admissible designs.*

Proof. (i) For an admissible design $\xi \in \Xi$ define

$$g(x) = \int K(u,x)f(u)\xi(\mathrm{d}u).$$

Represent this function in the form $g(x) = \lambda f(x) + h(x)$, where

$$\lambda = \frac{\int f(x)g(x)\xi(\mathrm{d}x)}{\int f^2(x)\xi(\mathrm{d}x)} \quad \text{and} \quad \int h(x)f(x)\xi(\mathrm{d}x) = 0;$$

that is, λf is the projection of g onto the line $\{cf(x) \mid c \in \mathbb{R}\}$ and h is the orthogonal complement in the space $L_2(\mathcal{X},\xi)$. Observing the definition of g we have

$$\begin{aligned} b(x,\xi) &= \frac{f(x)\int K(u,x)f(u)\xi(\mathrm{d}u)}{\iint K(u,v)f(u)f(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)} = \frac{f(x)g(x)}{\int f(x)g(x)\xi(\mathrm{d}x)} \\ &= \frac{\lambda f^2(x)+f(x)h(x)}{\lambda\int f^2(u)\xi(\mathrm{d}u)} = \frac{f^2(x)}{\int f^2(u)\xi(\mathrm{d}u)} + \frac{f(x)h(x)}{\lambda\int f^2(u)\xi(\mathrm{d}u)} \\ &= d(x,\xi) + \frac{f(x)h(x)}{\lambda\int f^2(u)\xi(\mathrm{d}u)} \end{aligned}$$

and therefore

$$b(x,\xi) - d(x,\xi) = \frac{f(x)h(x)}{\lambda \int f^2(u)\xi(\mathrm{d}u)} \,. \tag{4.7}$$

If the design ξ is optimal then, according to Theorem 3.1, it satisfies the condition $b(x,\xi) = d(x,\xi)$ for ξ-almost all $x \in \mathcal{X}$. Since the design ξ is admissible, it follows $f(x) \neq 0$ for ξ-almost all $x \in \mathcal{X}$, which yields that for the optimal design ξ we have $h(x) = 0$ for ξ-almost all $x \in \mathcal{X}$. This is equivalent to the statement (i).

(ii) If the condition (4.6) holds for all $x \in \mathcal{X}$, then in view of (4.7) we have $b(x,\xi) = d(x,\xi)$ for all $x \in \mathcal{X}$, and the optimality condition of Theorem 3.1 is satisfied.

(iii) If there is only one pair (ξ, λ) such that the condition (4.6) is satisfied for ξ-almost all $x \in \mathcal{X}$, then there is only one admissible design such that the necessary condition of optimality is satisfied. Hence the necessary condition becomes sufficient as well. □

4.3. *Optimality of the arcsine design.* In the remaining part of this section we will concentrate on the design problem for the domain $\mathcal{X} = [-1,1]$. To present results on optimality of the arcsine design for a number of one-parameter models, we start with the following lemma, which results in the theory of Fredholm-Volterra integral equations [see Mason and Handscomb (2002), Ch. 9, page 211].

Lemma 4.1 *The Chebyshev polynomials of the first kind $T_n(x) = \cos(n \arccos x)$ are the eigenfunctions of the integral operator with the kernel $H(x,v) = -\ln(x-v)^2/\sqrt{1-v^2}$. More precisely, for all $n = 0,1,\ldots$ we have for all $n \in \mathbb{N}$*

$$\lambda_n T_n(x) = -\int_{-1}^{1} T_n(v) \ln(x-v)^2 \frac{\mathrm{d}v}{\pi\sqrt{1-v^2}} \,, \quad x \in [-1,1],$$

where $\lambda_0 = 2\ln 2$ and $\lambda_n = 2/n$ for $n \geq 1$.

In the following theorem we give a new characterization of the arcsine distribution and present a number of models for which the arcsine distribution is the optimal design for the regression model (4.1) with correlation function $\gamma - \beta \ln(u-v)^2$, $\beta > 0, \gamma \geq 0$.

Theorem 4.2
(a) Let ζ be a random variable supported on the interval $[-1,1]$. Then ζ is given by the arcsine distribution with density (3.9) *if and only if the equality*

$$\mathbb{E}\, T_n(\zeta)\big(-\ln(\zeta - x)^2\big) = c_n T_n(x)$$

holds for almost all $x \in [-1, 1]$*, where* $c_n = 2/n$ *if* $n \in \mathbb{N}$ *and* $c_0 = 2 \ln 2$ *if* $n = 0$.
(b) Consider the one-parameter model (4.1) *with* $f(x) = T_n(x)$ *and the covariance kernel* $K(u, v) = \rho(u - v) = \gamma - \beta \ln(u - v)^2$ *with* $\beta > 0$ *and* $\gamma \geq 0$*. Then the optimal design is unique (in the class of admissible designs) and given by the arcsine distribution with density* (3.9).

Proof. For a proof of part (a) we notice that the part "if" of the statement follows from Lemma 4.1 and we should prove the part "only if". Nevertheless, we provide a proof of the part "if" since it will be the base for proving the part "only if".

Since the statement for $n = 0$ is proved in Schmidt and Zhigljavsky (2009), we consider the case $n \in \mathbb{N}$ in the rest of proof. Using the transformation $\varphi = \arccos u$ and $\psi = \arccos x$ we obtain $T_n(\cos \varphi) = \cos(n\varphi)$ and

$$\int_{-1}^{1} \frac{\ln(u - x)^2}{\pi\sqrt{1 - u^2}} T_n(u) \mathrm{d}u = \int_0^{\pi} \frac{\ln(\cos \varphi - x)^2}{\pi \sin \varphi} \cos(n\varphi) \sin \varphi \, \mathrm{d}\varphi.$$

Consequently, in order to prove part (a) of Theorem 4.2 we have to show that the function

$$\int_0^{\pi} \ln(\cos \varphi - \cos \psi)^2 \cos(n\varphi) \, \mu(\mathrm{d}\varphi)$$

is proportional to $\cos(n\psi)$ if and only if μ has a uniform density on the interval $[0, \pi]$. Extending μ to the interval $[0, 2\pi]$ as a symmetric (with respect to the center π) measure, $\mu(A) = \mu(2\pi - A)$, and defining the measure $\tilde{\mu}$ as $\tilde{\mu}(A) = \mu(2A)/2$ for all Borel sets $A \in [0, \pi]$, we obtain

$$\begin{aligned}
\int_0^{\pi} \ln(\cos \varphi - \cos \psi)^2 \cos(n\varphi) \, \mu(\mathrm{d}\varphi) &= \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln(\cos \varphi - \cos \psi)^2 \, \mu(\mathrm{d}\varphi) \\
&= \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln \Big(2 \sin \frac{\varphi - \psi}{2} \sin \frac{\varphi + \psi}{2} \Big)^2 \mu(\mathrm{d}\varphi) \\
&= \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln 2^2 \mu(\mathrm{d}\varphi) + \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln \Big(\sin \frac{\varphi - \psi}{2} \Big)^2 \mu(\mathrm{d}\varphi) \\
&\quad + \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln \Big(\sin \frac{\varphi + \psi}{2} \Big)^2 \mu(\mathrm{d}\varphi) \\
&= 0 + \int_0^{\pi} \cos(2n\varphi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(\mathrm{d}\varphi) + \int_0^{\pi} \cos(2n\varphi) \ln \sin^2(\varphi + \psi/2) \tilde{\mu}(\mathrm{d}\varphi) \\
&= 2 \int_0^{\pi} \cos(2n\varphi - n\psi + n\psi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(\mathrm{d}\varphi) \\
&= 2 \cos(n\psi) \int_0^{\pi} \cos(2n\varphi - n\psi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(\mathrm{d}\varphi) \\
&\quad + 2 \sin(n\psi) \int_0^{\pi} \sin(2n\varphi - n\psi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(\mathrm{d}\varphi).
\end{aligned}$$

The part "if" follows from the facts that the functions $\cos(2nz)\ln\sin^2(z)$ and $\sin(2nz)\ln\sin^2(z)$ are π-periodic and

$$\int_0^\pi \sin(2n\varphi - n\psi)\ln\sin^2(\varphi-\psi/2)\frac{\mathrm{d}\varphi}{\pi} = \int_0^\pi \sin(2n\varphi)\ln\sin^2(\varphi)\frac{\mathrm{d}\varphi}{\pi} = 0,$$

$$\int_0^\pi \cos(2n\varphi - n\psi)\ln\sin^2(\varphi-\psi/2)\frac{\mathrm{d}\varphi}{\pi} = \int_0^\pi \cos(2n\varphi)\ln\sin^2(\varphi)\frac{\mathrm{d}\varphi}{\pi} = -1/n.$$

To prove the part "only if", we need to show that the convolution of $\cos(2nz)\ln\sin^2(z)$ and $\tilde{\mu}(z)$, i.e.

$$\int_0^\pi \cos(2n(\varphi-t))\ln\sin^2(\varphi-t)\tilde{\mu}(\mathrm{d}\varphi),$$

is constant for almost all $t\in[0,\pi]$ if and only if $\tilde{\mu}$ is uniform; and the same holds for the convolution of $\sin(2nz)\ln\sin^2(z)$ and $\tilde{\mu}(z)$. This, however, follows from (Schmidt and Zhigljavsky, 2009, Lem. 3) since $\cos(2nz)\ln\sin^2(z)\in L^2([0,\pi])$ and all complex Fourier coefficients of these functions are non-zero. Indeed,

$$\int_0^\pi \cos(2nt)\ln\sin^2(t)\sin(2kt)\mathrm{d}t = 0 \quad \forall\ k\in\mathbb{Z}$$

$$\int_0^\pi \cos(2nt)\ln\sin^2(t)\cos(2kt)\mathrm{d}t = (\gamma_{|n+k|}+\gamma_{|n-k|})/2 \quad \forall\ k\in\mathbb{Z}\,,$$

where $\gamma_0 = -2\pi\log 2$ and $\gamma_k = -\pi/k$ for $k\in\mathbb{N}$, see formula 4.384.3 in Gradshteyn and Ryzhik (1965).

In order to prove part (b) of Theorem 4.2 we assume without loss of generality that $\beta = 1$. Next, assume also $\gamma = 0$ and thus consider the logarithmic kernel $\rho(x) = -\ln x^2$. Let p denote the density of the arcsine distribution defined in (3.9). Applying Lemma 4.1, we obtain that

$$\begin{aligned} B(\xi^*,\xi^*) &= \int_{-1}^1\int_{-1}^1 \rho(u-v)T_n(u)T_n(v)p(u)p(v)\mathrm{d}u\mathrm{d}v \\ &= \int_{-1}^1 T_n(u)\lambda T_n(u)\mathrm{d}u = \lambda M(\xi^*). \end{aligned}$$

Consequently, we obtain $b(x,\xi^*) = T_n(x)(\lambda_n M(\xi^*))^{-1}\lambda_n T_n(x)$ and

$$d(x,\xi^*) = M^{-1}(\xi^*)T_n^2(x).$$

Therefore the inequality (3.6) holds as an equality and the required statement (for $\gamma = 0$) immediately follows from Theorem 4.1.

To prove the statement for the function $\rho(x) = -\ln x^2 + \gamma$ with positive γ, we consider two cases: $n \in \mathbb{N}$ and $n = 0$. For $n \in \mathbb{N}$ we obtain

$$\int_{-1}^{1} (-\ln(u-x)^2 + \gamma) T_n(u) p(u) \mathrm{d}u = -\int_{-1}^{1} \ln(u-x)^2 T_n(u) p(u) \mathrm{d}u = \lambda_n T_n(x)$$

since $\int T_n(u)p(u)\mathrm{d}u = 0$ whenever $n \geq 1$. Consequently, $b(x,\xi^*) = \nu^{-1} T_n^2(x)$ which implies $b(x,\xi^*) = d(x,\xi^*)$ and the statement follows from Theorem 4.1.

For $n = 0$ we have $f(x) = T_0(x) \equiv 1$. It is easy to see that $M(\xi^*) = 1$ and $d(x,\xi^*) = 1$. We also have $B(\xi^*,\xi^*) = \gamma - \lambda_0$,

$$B(\xi^*,\xi_x) = \int_{-1}^{1} (-\ln(u-x)^2 + \gamma) p(u) \mathrm{d}u = \gamma - \int_{-1}^{1} \ln(u-x)^2 p(u) \mathrm{d}u = \gamma - \lambda_0$$

and, consequently, $b(x,\xi^*) = B(\xi^*,\xi_x)/B(\xi^*,\xi^*) = 1$. Thus, $d(x,\xi^*) = b(x,\xi^*)$ for all $x \in [-1,1]$ and the statement again follows from Theorem 4.1. □

4.4. *Generalized arcsine designs.* Take any $\alpha \in (0,1)$ and consider the Gegenbauer polynomials $C_m^{(\alpha)}(x)$ which are orthogonal with respect to the weight function

$$p_\alpha(x) = \frac{(\Gamma(\alpha + \frac{1}{2}))^2}{2^\alpha \Gamma(2^{\alpha+1})} (1-x^2)^{\alpha - 1/2}, \quad x \in [-1,1]. \tag{4.8}$$

For the choice $\alpha = 0$ the Gegenbauer polynomials $C_m^{(\alpha)}(x)$ are proportional to the Chebyshev polynomials $T_m(x)$. Throughout this paper we will call the corresponding beta-distributions *generalized arcsine designs* emphasizing the fact that the distribution is symmetric and the parameter α varies in the interval $(0,1)$. The following result establishes an analogue of Lemma 4.1 for the kernel

$$H(u,v) = \frac{1}{|u-v|^\alpha (1-v^2)^{(1-\alpha)/2}}. \tag{4.9}$$

It appears in the theory of Fredholm-Volterra integral equations of the first kind with special kernel, see Fahmy et al. (1999).

Lemma 4.2 *The Gegenbauer polynomials $C_n^{(\alpha/2)}(x)$ are the eigenfunctions of the integral operator with the kernel defined in* (4.9)*. More precisely, for* *all* $n = 0, 1, \ldots$ *we have*

$$\lambda_n C_n^{(\alpha/2)}(x) = -\int_{-1}^{1} \frac{1}{|x-v|^\alpha} C_n^{(\alpha/2)}(v) \frac{\mathrm{d}v}{(1-v^2)^{(1-\alpha)/2}}$$

for all $x \in [-1, 1]$*, where* $\lambda_n = \frac{\pi\Gamma(n+\alpha)}{\cos(\alpha\pi/2)\Gamma(\alpha)n!}$.

In the following result we generalize Theorem 8 of Zhigljavsky et al. (2010) from the case of location scale model to more general one-parameter models.

Theorem 4.3 *Consider the one-parameter model* (4.1) *with* $f(x) = C_n^{(\frac{\alpha}{2})}(x)$ *and the covariance kernel* $K(u, v) = \rho(u - v) = \gamma + \beta/|u - v|^\alpha$*, where* $0 \leq \alpha < 1$*,* $\gamma \geq 0$*,* $\beta > 0$ *and* $n = 0, 1, \ldots$ *Then the design with the generalized arcsine density*

$$p_{\alpha/2}(x) = \frac{2^{-\alpha}}{B((1+\alpha)/2, (1+\alpha)/2)} (1 - x^2)^{(\alpha-1)/2}$$

satisfies the necessary condition of optimality stated in Theorem 4.1.

Proof. It is easy to see that the optimal design does not depend on β and we thus assume $\beta = 1$. The statement for $\gamma = 0$ follows from Lemma 4.2.

To prove the statement for the kernel $\rho(x) = 1/|x|^\alpha + \gamma$ with positive γ we recall the definition of p_α in (4.8) and we consider the two cases $n \in \mathbb{N}$ and $n = 0$ separately. For $n \in \mathbb{N}$ we have by Lemma 4.2

$$\int \Big(\frac{1}{|u-x|^\alpha} + \gamma\Big) C_n^{(\frac{\alpha}{2})}(u) p_{\frac{\alpha}{2}}(u)\mathrm{d}u = \int \frac{1}{|u-x|^\alpha} C_n^{(\frac{\alpha}{2})}(u) p_{\frac{\alpha}{2}}(u)\mathrm{d}u \propto C_n^{(\frac{\alpha}{2})}(x)$$

since $\int C_n^{(\alpha/2)}(u) p_{\alpha/2}(u)\mathrm{d}u = 0$. Then $d(x, \xi^*) = (C_n^{(\alpha/2)}(x))^2 M^{-1}(\xi^*) = b(x, \xi^*)$ and ξ^* obviously satisfies the necessary condition in Theorem 4.1.

For $n = 0$ we have that $M(\xi^*) = 1$, $B(\xi^*, \xi^*) = \gamma + \lambda_\alpha$ for some constant λ_α and

$$B(\xi^*, \xi_x) = \int \Big(\frac{1}{|u-x|^\alpha} + \gamma\Big) p_{\alpha/2}(u)\mathrm{d}u = \gamma + \int \frac{1}{|u-x|^\alpha} p_{\alpha/2}(u)\mathrm{d}u = \gamma + \lambda_\alpha.$$

Consequently, $d(x, \xi^*) = b(x, \xi^*) = 1$ which completes the proof. □

5. Universally optimal designs for multi-parameter models. In this section we consider the matrix $D(\xi)$ defined in (2.6) as the matrix optimality criterion which we are going to minimize on the set of all admissible designs, which is the set Ξ of all probability measures $\xi \in \Xi$ supported on the design space $\mathcal{X}$ such that the matrices $B(\xi, \xi)$ and $M^{-1}(\xi)$ (and therefore the matrix $D(\xi)$) are well-defined. We call a design ξ^* universally optimal if $D(\xi^*) \le D(\xi)$ in the sense of Loewner ordering for any admissible design $\xi \in \Xi$.

Note that a design ξ^* is universally optimal if and only if ξ^* is c-optimal for any vector $c \neq 0$. Correspondingly, a necessary condition for the universal optimality of a design ξ^* is a combination of necessary conditions of c-optimality for all $c \neq 0$. Moreover, a design that is universally optimal is also D-, E-, A- and L-optimal for any matrix $L \ge 0$.

5.1. *A general result.*

Theorem 5.1 *Consider the regression model* (1.1) *with correlation structure* (2.1) *and a design* ξ^* *with support on* $\mathcal{X}$. *Assume that the regression functions* $f_1(x), \ldots, f_m(x)$ *are linearly independent eigenfunctions of the integral operator defined by the correlation kernel* K *with respect to the design* ξ^*, *that is*

$$\int K(x,u) f(u) \xi^*(\mathrm{d}u) = \Lambda f(x) \tag{5.1}$$

for all $x \in \mathcal{X}$, *where* Λ *is an* $m \times m$ *diagonal matrix. If the matrix* $M(\xi^*)$ *is non-degenerate then the design* ξ^* *satisfies the necessary condition for the universal optimality for the model* (1.1); *that is, the inequality*

$$f^T(x) M^{-1}(\xi^*) B(\xi^*,\xi^*) c c^T M^{-1}(\xi^*) f(x) \le c^T M^{-1}(\xi^*) B(\xi^*,\xi_x) M^{-1}(\xi^*) c \tag{5.2}$$

holds for all vectors $c \neq 0$;

Proof. The assumption (5.1) implies

$$B(\xi^*,\xi^*) \;=\; \int \left[\int K(u,x) f(u) \xi^*(\mathrm{d}u) \right] f^T(x) \xi^*(\mathrm{d}x) \;=\; \Lambda M(\xi^*)\,, \tag{5.3}$$

while the symmetry of the correlation kernel $K(u,x)$ and (5.1) yield

$$\int K(u,x) f^T(x) \xi^*(\mathrm{d}x) = f^T(u) \Lambda\,. \tag{5.4}$$

Similarly to (5.3) it follows using (5.4)

$$B(\xi^*,\xi^*) \;=\; \int f(u) \left[\int K(u,x) f^T(x) \xi^*(\mathrm{d}x) \right] \xi^*(\mathrm{d}u) \;=\; M(\xi^*) \Lambda\,. \tag{5.5}$$

Expressing $B(\xi,\xi)$ by the formula (5.5) in the expression (3.2) for the function φ we obtain

$$\begin{aligned}\varphi(x,\xi^*) &= f^T(x)M^{-1}(\xi^*)B(\xi^*,\xi^*)M^{-1}(\xi^*)\,CM^{-1}(\xi^*)f(x)\\ &= f^T(x)\Lambda M^{-1}(\xi^*)CM^{-1}(\xi^*)f(x)\,,\end{aligned}$$

where $C = cc^T$ for the c-optimality criterion determined by the vector c. For the function $b(x,\xi)$ defined in (3.3), we obtain using (5.1)

$$\begin{aligned}b(x,\xi^*) &= f^T(x)M^{-1}(\xi^*)CM^{-1}(\xi^*)\int K(u,x)f(u)\xi^*(\mathrm{d}u)\\ &= f^T(x)M^{-1}(\xi^*)CM^{-1}(\xi^*)\Lambda f(x)\,.\end{aligned}$$

In view of the identity $a^TAa = a^TA^Ta$, which holds for any square matrix A and vector a of suitable size, the last two formulas imply that identity $\varphi(x,\xi) = b(x,\xi)$ for all x. Therefore ξ^* satisfies the necessary condition for c-optimality, for any vector $c \neq 0$. □

The following result immediately follows from Theorem 5.1.

Corollary 5.1 *Under the assumptions of Theorem 5.1, the design ξ^* satisfies the necessary condition for the universal optimality for any submodel $y = \theta_{i_1}f_{i_1}(x)+\theta_{i_2}f_{i_2}(x)+\ldots+\theta_{i_k}f_{i_k}(x)+\varepsilon$ of the original regression model* (1.1)*, where $1 \le k < m$ and $1 \le i_1 < \ldots < i_k \le m$, as well as for any other model of the form $y = \sum_{i=1}^m \theta_i g_i(x) + \varepsilon$, where the vector of regression functions $g(x) = (g_1(x),\ldots,g_m(x))^T$ is a linear transformation of the vector of regression functions $f(x) = (f_1(x),\ldots,f_m(x))^T$, that is, $g(x) = Lf(x)$ with a non-degenerate $m \times m$ matrix L.*

5.2. *Optimality of the arcsine design.* The following theorem establishes universal optimality of the arcsine design for the polynomial regression model with the logarithmic correlation function. To our best knowledge this is the first explicit solution of an optimal design problem for a regression model with correlated observations with more than one parameter.

Theorem 5.2 *Consider the polynomial regression model* (1.1) *with $f(x) = (1,x,\ldots,x^{m-1})^T$, $x \in [-1,1]$, and covariance kernel $K(x,y) = -\ln|x-y|^2$. Then the design ξ_a with the arcsine density* (3.9) *is universally optimal.*

Proof. Let $c \neq 0$ be any fixed vector in $\mathbb{R}^m$. Consider the c-optimality criterion $\Psi(\xi) = c^TM^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)c$. Straightforward calculation shows that the arcsine design ξ_a satisfies the optimality condition for c-optimality. We will demonstrate that there is no other design that is better than ξ_a

with respect to the criterion $\Psi(\xi)$. Specifically, we will show that if there exists another design ξ_0 satisfying the necessary condition for c-optimality, then $\Psi(\xi_a) = \Psi(\xi_0)$. Because the vector c is arbitrary the assertion is then obvious. The proof of this equality is divided into three steps.

(i) For any design ξ_0, we have the following representation of the integral $\int K(u,x)f(u)\xi_0(\mathrm{d}u)$ in two orthogonal components:

$$\int K(u,x)f(u)\xi_0(\mathrm{d}u) = \Lambda_0 f(x) + g_0(x) \tag{5.6}$$

where Λ_0 is a $m{\times}m$-matrix and the function g_0 satisfies $\int g_0(x)f^T(x)\xi_0(\mathrm{d}x) = \mathbf{0} \in \mathbb{R}^m$. Using (5.6) we rewrite the function $b(x,\xi_0)$ in (3.3) as

$$\begin{aligned} b(x,\xi_0) &= \mathrm{tr}\left(M^{-1}(\xi_0)\, cc^T M^{-1}(\xi_0) \int K(u,x)f(u)f^T(x)\xi_0(\mathrm{d}u)\right) \\ &= c^T M^{-1}(\xi_0)\Lambda_0 f(x)f^T(x)M^{-1}(\xi_0)\, c + c^T M^{-1}(\xi_0)g_0(x)f^T(x)M^{-1}(\xi_0)c. \end{aligned}$$

Therefore, $b(x,\xi_0) = \varphi(x,\xi_0) + r(x,\xi_0)$, where the function φ is defined in (3.2) and

$$r(x,\xi_0) = c^T M^{-1}(\xi_0)g_0(x)f^T(x)M^{-1}(\xi_0)c.$$

Thus, the necessary condition for c-optimality of ξ_0 can be reformulated as follows: if the design ξ_0 is c-optimal then $r(x,\xi_0) \geq 0$ for all x and $r(x,\xi_0) = 0$ for $x \in \operatorname{supp}\xi_0$.

Suppose now that ξ_0 is a c-optimal design. Since $f^T(x)M^{-1}(\xi_0)c \neq 0$ for almost all $x \in [-1,1]$, we have $a^T g_0(x) = 0$ for all $x \in \operatorname{supp}\xi_0$, where $a^T = c^T M^{-1}(\xi_0)$.

(ii) Now we want to prove that $a^T g_0(x) = 0$ for all $x \in [-1,1]$. Note that elements of $B(\xi,\xi)$ are equal to infinity and $\Psi(\xi) = \infty$ if ξ has a discrete part. Therefore, ξ_0 is a continuous measure or a singular continuous measure or a mixture of them. This implies that $\operatorname{supp}\xi_0$ contains more than a countable number of points.

From the optimality condition (or multiplying (5.6) by a^T on the left) it follows

$$a^T \int_{-1}^{1} K(u,x)f(u)\xi_0(\mathrm{d}u) = a^T \Lambda_0 f(x)$$

for all $x \in \operatorname{supp}\xi_0$. For the assumed regression model, we have

$$\sum_{j=0}^{m-1} a_j \int_{-1}^{1} \ln(x-u)^2 T_j(u)\xi_0(\mathrm{d}u) = \sum_{j=0}^{m-1} b_j T_j(x) \tag{5.7}$$

for all $x \in \operatorname{supp} \xi_0$, where $b^T = (b_0, \ldots, b_{m-1}) = a^T \Lambda_0$.

Following the arguments given in the proof of Theorem 4.2, we can rewrite (5.7) as

$$\sum_{j=0}^{m-1} 2a_j \int_0^\pi \ln \sin^2(\varphi - \psi/2) \cos(2j\varphi) \tilde{\mu}(\mathrm{d}\varphi) = \sum_{j=0}^{m-1} b_j \cos(2j\psi/2) \tag{5.8}$$

where $\psi \in \Omega = \{\psi : \cos(\psi) \in \operatorname{supp} \xi_0\}$ and the measure $\tilde{\mu}$ denotes the "symmetrization" of the design ξ_0 introduced in the proof of Theorem 4.2. Note that the left hand side of (5.8) is a convolution. Thus, the equation (5.8) can be written as

$$(K_s \star \nu)(v) = g(v) \tag{5.9}$$

where $v = \psi/2, 2v \in \Omega$, $K_s(v) = \ln \sin^2(v)$, $\nu(\mathrm{d}\varphi) = \sum_{j=0}^{m-1} a_j \cos(2j\varphi)\tilde{\mu}(\mathrm{d}\varphi)$ and $g(v) = \sum_{j=0}^{m-1} b_j \cos(2jv)$.

Note that the function $G(v) := (K_s \star \nu)(v)$ is π-periodic and continuous and, therefore, can be uniformly approximated by a linear combination of $\{e^{2ikv} | k \in \mathbb{Z}\}$. Consequently, we have the equality $G(v) = g(v)$ for all v such that $2v \in \Omega$ and the fact that $g(v)$ is a finite sum of cosines implies that $G(v) = g(v)$ for all v. Thus, we have proved that $a^T g_0(x) = 0$ for all $x \in [-1, 1]$.

(iii) We are now ready to prove that $\Psi(\xi_a) = \Psi(\xi_0)$. By the definition of $B(\xi_0, \xi_a)$ and (5.6) we obtain

$$B(\xi_0, \xi_a) = \Lambda_0 M(\xi_a) + \int g_0(x) f^T(x) \xi_a(\mathrm{d}x) \tag{5.10}$$

and Lemma 4.1 implies $B(\xi_a, \xi_0) = \Lambda_a M(\xi_0)$, $B(\xi_a, \xi_a) = \Lambda_a M(\xi_a)$ for some matrix $\Lambda_a \in \mathbb{R}^{m \times m}$ Since $B(\xi_0, \xi_a) = B^T(\xi_a, \xi_0)$, we obtain another representation for $B(\xi_0, \xi_a)$:

$$B(\xi_0, \xi_a) = M(\xi_0) \Lambda_a^T. \tag{5.11}$$

Multiplying (5.10) and (5.11) by a^T on the left and using the identity $a^T g_0(x) = 0$ for all $x \in [-1, 1]$, we get

$$a^T B(\xi_0, \xi_a) = a^T \Lambda_0 M(\xi_a) + \int a^T g_0(x) f^T(x) \xi_a(\mathrm{d}x) = a^T \Lambda_0 M(\xi_a) \tag{5.12}$$

and

$$a^T B(\xi_0, \xi_a) = a^T M(\xi_0) \Lambda_a^T. \tag{5.13}$$

Combining (5.12) and (5.13) we obtain

$$a^T \Lambda_0 M(\xi_a) = c^T \Lambda_a^T. \tag{5.14}$$

Let us now compute $\Psi(\xi_a)$ and $\Psi(\xi_0)$. Using (5.14), we get

$$\begin{aligned}\Psi(\xi_a) &= c^T D(\xi_a)c = c^T M^{-1}(\xi_a)B(\xi_a,\xi_a)M^{-1}(\xi_a)c \\ &= c^T M^{-1}(\xi_a)\Lambda_a M(\xi_a)M^{-1}(\xi_a)c = c^T M^{-1}(\xi_a)\Lambda_a c \\ &= c^T \Lambda_a^T M^{-1}(\xi_a)c = a^T \Lambda_0 M(\xi_a)M^{-1}(\xi_a)c = a^T \Lambda_0 c.\end{aligned}$$

Using $\Lambda_0 = B(\xi_0,\xi_0)M^{-1}(\xi_0)$ (which also follows from (5.6) and part (ii) of this proof) and the definition of a, we get

$$\Psi(\xi_0) = c^T D(\xi_0)c = c^T M^{-1}(\xi_0)B(\xi_0,\xi_0)M^{-1}(\xi_0)c = a^T \Lambda_0 c.$$

Comparing the last two formulas we obtain the desired equality $\Psi(\xi_a) = \Psi(\xi_0)$, which completes the proof of Theorem 5.2. □

The next result follows from Theorems 5.2 and 4.2.

Corollary 5.2 *The statement of Theorem 5.2 remains true for the covariance kernel* $K(x,y) = \gamma - \beta \ln |x-y|^2$ *with* $\gamma \geq 0$, $\beta > 0$.

5.3. *Optimality and the generalized arcsine design.* The following result for the generalized arcsine design is a direct consequence of Theorems 4.3 and 5.1.

Corollary 5.3 *Consider the polynomial regression model* (1.1) *with* $f(x) = (1, x, x^2, \dots, x^{m-1})^T$, $x \in [-1,1]$, *and covariance kernel* $K(x,y) = \gamma + \beta/|x-y|^\alpha$ *with* $\alpha \in (0,1)$, $\gamma \geq 0$, $\beta > 0$. *Then the design with generalized arcsine density defined in* (4.8) *satisfies the necessary conditions for universal optimality.*

5.4. *Optimality and Mercer's theorem.* In this section we consider the case when the regression functions are proportional to eigenfunctions from Mercer's theorem. To be precise let $\mathcal{X}$ denote a compact subset of a metric space and let ν denote an absolute continuous probability measure on the corresponding Borel field with positive density. Consider the integral operator

$$T_K(f)(\cdot) = \int_{\mathcal{X}} K(\cdot, y) f(y) \nu(dy) \tag{5.15}$$

on $L_2(\nu)$. Under certain assumptions on the kernel (for example if K is symmetric, continuous and positive definite) T_K defines a symmetric, compact self-adjoint operator. In this case Mercer's Theorem [see e.g. Kanwal (1997)] shows that there exist a countable number of eigenfunctions $\varphi_1, \varphi_2, \ldots$ with positive eigenvalues $\lambda_1, \lambda_2, \ldots$ of the operator K, that is

$$T_k(\varphi_\ell) = \lambda_\ell \varphi_\ell \ , \ \ell = 1, 2, \ldots \tag{5.16}$$

Theorem 5.3 *Let $\mathcal{X}$ be a compact subset of a metric space and assume that the covariance kernel $K(x,u)$ defines a integral operator T_K of the form* (5.15)*, where the eigenfunctions satisfy* (5.16) *and the measure ν is absolute continuous with positive density. Consider the regression model* (1.1) *with $f(x) = L(\varphi_{i_1}(x), \ldots, \varphi_{i_m}(x))^T$ and the covariance kernel $K(x,u)$, where $L \in \mathbb{R}^{m\times m}$ is a non-singular matrix. Then the design ν satisfies the necessary conditions for universal optimality.*

Proof. By construction, the regression functions are the eigenfunctions of the integral operator with covariance kernel $K(x,u)$. Therefore the result follows by exactly the same arguments for the measure ν on $\mathcal{X}$ as given in the proof of Theorem 5.1. □

We note that the Mercer expansion is known analytically for certain covariance kernels. For example, if ν is the uniform distribution on the interval $\mathcal{X} = [-1,1]$ and the covariance kernel is of exponential type, that is $K(x,u) = e^{-\lambda|x-u|}$, then the eigenfunctions are given by

$$\varphi_k(x) = \sin(\omega_k x + k\pi/2), \quad k \in \mathbb{N},$$

where $\omega_1, \omega_2, \ldots$ are positive roots of the equation $\tan(2\omega) = -2\lambda\omega/(\lambda^2 - \omega^2)$. Similarly, consider as a second example, the covariance kernel $K(x,u) = \min\{x,u\}$ and $\mathcal{X} = [0,1]$, In this case the eigenfunctions of the corresponding integral operator are given by

$$\varphi_k(x) = \sin((k+1/2)\pi x), \quad k \in \mathbb{N}.$$

In the following subsection we provide a further example of the application of Mercer's theorem.

5.5. *Uniform design for periodic covariance functions.* Consider the regression functions

$$f_j(x) = \begin{cases} 1 & \text{if} \quad j = 1 \\ \sqrt{2}\cos(2\pi(j-1)x) & \text{if} \quad j \geq 2 \end{cases} \tag{5.17}$$

and the design space $\mathcal{X} = [0,1]$. Assume that the correlation function $\rho(x)$ is periodic with period 1, that is $\rho(x) = \rho(x+1)$, and let a correlation kernel be defined by $K(u,v) = \rho(u-v)$. An example of the covariance kernel $\rho(x)$ satisfying this property is provided by a convex combination of the functions $\{\cos(2\pi x), \cos^2(2\pi x), \ldots\}$.

Theorem 5.4 *Consider the regression model* (1.1) *with regression functions* $f_j(x)$ *defined in* (5.17) *and a correlation function* $\rho(x)$ *that is periodic with period* 1. *Then the uniform design satisfies the necessary conditions for universal optimality.*

Proof. We will show that the identity

$$\int_0^1 K(u,x) f_j(u) \mathrm{d}u = \int_0^1 \rho(u-x) f_j(u) \mathrm{d}u = \lambda_j f_j(x) \tag{5.18}$$

holds for all $x \in [0,1]$, where $\lambda_j = \int \rho(u) f_j(u) \mathrm{d}u$ $(j \geq 1)$. To prove (5.18), we define $A_j(v) = \int_0^1 \rho(u-v) f_j(u) \mathrm{d}u$ which should be shown to be $\lambda_j f_j(x)$. For $j = 1$ we have $A_1(v) = \lambda_1$ because $\int_0^1 \rho(u-v)\mathrm{d}u = \int_0^1 \rho(u)\mathrm{d}u = \lambda_1$ by the periodicity of the function $\rho(x)$. For $j = 2, 3, \ldots$ we note that

$$\begin{aligned} A_j(v) &= \int_0^1 \rho(u-v) f_j(u) \mathrm{d}u = \int_{-v}^{1-v} f_j(u+v) \rho(u) \mathrm{d}u \\ &= \int_0^{1-v} f_j(u+v)\rho(u)\mathrm{d}u + \int_{-v}^0 f_j(u+v)\rho(u)\mathrm{d}u. \end{aligned}$$

Because of the periodicity we have

$$\int_{-v}^0 f_j(u+v)\rho(u)\mathrm{d}u = \int_{1-v}^1 f_j(u+v)\rho(u)\mathrm{d}u$$

which gives $A_j(v) = \int_0^1 f_j(u+v)\rho(u)\mathrm{d}u$. A simple calculation now shows

$$A_j''(v) = -b_j^2 A_j(v) \tag{5.19}$$

where $b_j^2 = (2\pi(j-1))^2$ and

$$\begin{aligned} A_j(0) &= \int_0^1 \cos(2\pi(j-1)u)\rho(u)\mathrm{d}u = \lambda_j \\ A_j'(0) &= -b_j \int_0^1 \sin(2\pi(j-1)u)\rho(u)\mathrm{d}u = 0. \end{aligned}$$

Therefore (from the theory of differential equations) the unique solution of (5.19) is of the form $A_j(v) = c_1 \cos(b_j v) + c_2 \sin(b_j v)$, where c_1 and c_2 are determined by initial conditions, that is $A(0) = c_1 = \lambda_j, A'(0) = b_j c_2 = 0$. This yields $A_j(v) = \lambda_j \cos(2\pi(j-1)v) = \lambda_j f_j(v)$ and proves the identity (5.18). □

5.6. *Optimality for the triangular covariance function.* Let us now consider the triangular correlation function defined by

$$\rho(x) = \max\{0, 1 - \lambda|x|\}. \tag{5.20}$$

The following theorem presents a candidate for the optimal design in the linear model with a triangular correlation function.

Theorem 5.5 *Consider the model* (1.1) *with* $f(x) = (1, x)^T$, $\mathcal{X} = [-1, 1]$, *and the triangular correlation function* (5.20).

(a) If $\lambda \in (0, 1/2]$, *then the design* $\xi^* = \{-1, 1; 1/2, 1/2\}$ *satisfies the necessary condition for universal optimality.*

(b) If $\lambda \in \mathbb{N}$, *then the design supported at* $2\lambda+1$ *points* $x_k = -1+k/\lambda$, $k = 0, 1, \ldots, 2\lambda$, *with equal weights satisfies the necessary condition for universal optimality.*

Proof. For a proof of part (a) we use arguments as given in the proof of Theorem 4.2 in Zhigljavsky et al. (2010) and obtain $\int \rho(x-u) f_i(u) \xi^*(du) = f_i(x)$ for $i = 1, 2$. Thus, the assumptions of Theorem 5.1 are fulfilled.

Part (b). Straightforward but tedious calculations show that $M(\xi^*) = \mathrm{diag}(1, \gamma)$, where $\gamma = \sum_{k=0}^{2\lambda+1} x_k^2/(2\lambda + 1) = (\lambda + 1)/(2\lambda)$. Also we have $\int \rho(x-u) f_i(u) \xi^*(du) = f_i(x)$ for $i = 1, 2$. Thus, the assumptions of Theorem 5.1 are fulfilled. □

Note that the designs provided in Theorem 5.5 are optimal for the location scale model, see Zhigljavsky et al. (2010). It is also worthwhile to note that unlike the results of previous subsections the result of Theorem 5.5 cannot be extended to polynomial models of higher order.

6. Numerical construction of optimal designs.

6.1. *An algorithm for computing optimal designs for non-singular kernels.* Numerical computation of optimal designs for a common linear regression model (1.1) with given correlation function can be performed by an extension of the multiplicative algorithm proposed by Dette et al. (2008b). Note that the results of this algorithm is a discrete design which approximates the optimal design with arbitrary precision.

To be precise, let $\xi^{(r)} = \{x_1, \ldots, x_n; w_1^{(r)}, \ldots, w_n^{(r)}\}$ be a design at the iteration r. Assume that $x_1, \ldots, x_n$ is a rather uniform dense set in the interval $[-1, 1]$ and $w_1^{(0)}, \ldots, w_n^{(0)}$ are nonzero weights, for example, uniform.

We propose the following updating rule for the weights

$$w_i^{(r+1)} = \frac{w_i^{(r)}\big(\psi(x_i,\xi^{(r)})-\beta\big)}{\sum_{j=1}^n w_j^{(r)}\big(\psi(x_j,\xi^{(r)})-\beta\big)} \qquad i=1,\dots,n, \tag{6.1}$$

where β is a tuning parameter, $\psi(x,\xi)=\varphi(x,\xi)/b(x,\xi)$ and the functions $\varphi(x,\xi)$ and $b(x,\xi)$ are defined in (3.2) and (3.3), respectively. The condition (3.5) takes the form $\psi(x,\xi^*)\le 1$. Note that $\psi(x,\xi)\ge 0$ for all x and ξ. The rule (6.1) means that the weight of a point increases if the condition (3.5) does not hold.

The algorithm above can be easily extended to cover the case of singular covariance kernels. Alternatively, a singular kernel can be approximated by a non-singular one using the technique described in Zhigljavsky et al. (2010), Section 4.

6.2. *Efficiencies of the uniform and arcsine densities.* In the present section we numerically study the D-efficiency of two designs for different models. Specifically, we consider the uniform design and the arcsine design for the model (1.1) with $f(x)=(1,x,\dots,x^{m-1})^T$ and different correlation functions where the design space is given by the interval $[-1,1]$. We determine the D-efficiency as

$$\text{Eff}(\xi)=\left(\frac{\det D(\xi^*)}{\det D(\xi)}\right)^{1/m},$$

where ξ^* is the design computed by the algorithm described in the previous section. We considered polynomial regression models of degree ≤ 3 and the correlation functions $\rho(x)=e^{-\lambda|x|}$ and $\rho(x)=e^{-\lambda x^2}$ for various values of λ. The results are depicted in Tables 1 and 2, respectively. We observe that the efficiency of the arcsine design is always larger than the efficiency of the uniform design. Moreover, the absolute difference between the efficiencies of the two designs increases as the degrees m of the polynomial increases. On the other hand, the efficiency of the uniform design and the arcsine design decreases as m increases.

7. Conclusions. In this paper we have addressed the problem constructing optimal designs for least squares estimation in regression models with correlated observations. The main challenge in problems of this type is that - in contrast to "classical" optimal design theory for uncorrelated data - the corresponding optimality criteria are not convex (except for the location

TABLE 1
D-Efficiencies of the uniform design ξ_u and the arcsine design ξ_a for the polynomial regression model of degree $m-1$ and the exponential correlation function $\rho(x) = e^{-\lambda|x|}$.

	λ	0.5	1.5	2.5	3.5	4.5	5.5
$m = 1$	Eff(ξ_u)	0.913	0.888	0.903	0.919	0.933	0.944
	Eff(ξ_a)	0.966	0.979	0.987	0.980	0.968	0.954
$m = 2$	Eff(ξ_u)	0.857	0.832	0.847	0.867	0.886	0.901
	Eff(ξ_a)	0.942	0.954	0.970	0.975	0.973	0.966
$m = 3$	Eff(ξ_u)	0.832	0.816	0.826	0.842	0.860	0.876
	Eff(ξ_a)	0.934	0.938	0.954	0.968	0.976	0.981
$m = 4$	Eff(ξ_u)	0.826	0.818	0.823	0.835	0.849	0.864
	Eff(ξ_a)	0.934	0.936	0.945	0.957	0.967	0.975

TABLE 2
D-Efficiencies of the uniform design ξ_u and the arcsine design ξ_a for the polynomial regression model of degree $m-1$ and the Gaussian correlation function $\rho(x) = e^{-\lambda x^2}$.

	λ	0.5	1.5	2.5	3.5	4.5	5.5
$m = 1$	Eff(ξ_u)	0.758	0.789	0.811	0.830	0.842	0.853
	Eff(ξ_a)	0.841	0.907	0.924	0.932	0.934	0.935
$m = 2$	Eff(ξ_u)	0.756	0.698	0.709	0.725	0.739	0.753
	Eff(ξ_a)	0.843	0.833	0.853	0.868	0.877	0.885
$m = 3$	Eff(ξ_u)	0.803	0.662	0.684	0.699	0.711	0.720
	Eff(ξ_a)	0.866	0.771	0.818	0.844	0.859	0.869
$m = 4$	Eff(ξ_u)	0.797	0.630	0.617	0.627	0.648	0.665
	Eff(ξ_a)	0.842	0.713	0.722	0.746	0.776	0.799

scale model). Necessary conditions for optimality have been derived, which can be easily used to identify candidates of optimal designs. By relating the design problem to an integral operator problem these candidates can be identified explicitly for a broad class of regression models and correlation structures. Moreover, for one parameter regression models these designs can be shown to be optimal in many cases. Particular attention is paid to the classical polynomial regression model with a logarithmic covariance kernel, where it is proved that the arcsine distribution is universally optimal (for any degree).

So far optimal designs for regression models with correlated observations have only be derived explicitly for the location scale model and to our best knowledge the results presented in this paper provide the first explicit solutions to this type of problem for a general class of models with one parameter and specific models with more than one parameter. By investigating more integral operator problems it is expected that further design problems can be solved explicitly in the future. It is usually easy to find designs explicitly satisfying the necessary condition of universal optimality. However, proving in fact optimality of the candidate designs is substantially harder and the necessary arguments will depend on specific properties of the model under investigation (see the arguments given in the proof of Theorem 5.2).

We finally point out that we have concentrated on the construction of optimal designs for least squares estimation (LSE), because the best linear unbiased estimator (BLUE) requires the knowledge of the correlation matrix. While the BLUE is often sensitive with respect to misspecification of the correlation structure the corresponding optimal designs for LSE show a remarkable robustness [see Dette et al. (2009)]. Moreover, the difference between BLUE and LSE is often surprisingly small and in many cases BLUE and LSE with certain correlation functions are asymptotically equivalent [see Rao (1967), Kruskal (1968)].

Indeed, consider the location scale model $y(x) = \theta + \varepsilon(x)$ with $K(u, v) = \rho(u - v)$, where the knowledge of a full trajectory of a process $y(x)$ is available. Define the (linear unbiased) estimate $\hat{\theta}(G) = \int y(x) \mathrm{d}G(x)$, where $G(x)$ is a distribution function of a signed probability measure. A remarkable result of Grenander (1950) states that the "estiamtor" $\hat{\theta}(G^*)$ is BLUE if and only if $\int \rho(u - x) \mathrm{d}G^*(u)$ is constant for all $x \in \mathcal{X}$. This result was extended by Näther (1985a), Sect. 4.3 to the case of random fields with constant mean. Consequently, if $G^*(x)$ is a distribution function of a non-signed (rather than signed) probability measure, then LSE coincides with BLUE and an asymptotic optimal design for LSE is also an asymptotic optimal design for BLUE. Hajek (1956) proved that $G^*(x)$ is a distribution function of a non-signed probability measure if the correlation function $\rho(x)$ is convex on the interval $(0, \infty)$. Zhigljavsky et al. (2010) showed that $G^*(x)$ is a proper distribution function for a certain families of correlation functions including non-convex ones. An interesting direction of future research is to investigate if and how these results can be extended to more general regression models as considered in this paper.

Acknowledgements. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG). Parts of this paper were written during a visit of the authors at the Isaac Newton Institute, Cambridge, UK, and the authors would like to thank the institute for its hospitality and financial support.

References.

Bickel, P. J. and Herzberg, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Annals of Statistics*, 7(1):77–95.

Bickel, P. J., Herzberg, A. M., and Schilling, M. F. (1981). Robustness of design against autocorrelation in time II: Optimality, theoretical and numerical results for the first-order autoregressive process. *Journal of the American Statistical Association*, 76(376):870–877.

Boltze, L. and Näther, W. (1982). On effective observation methods in regression models with correlated errors. *Math. Operationsforsch. Statist. Ser. Statist.*, 13:507–519.

Dette, H., Kunert, J., and Pepelyshev, A. (2008a). Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica*, 18(1):135–154.

Dette, H., Leonenko, N. N., Pepelyshev, A., and Zhigljavsky, A. (2009). Asymptotic optimal designs under long-range dependence error structure. *Bernoulli*, 15:1036–1056.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2008b). Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Computational Statistics and Data Analysis*, 53(2):312–320.

Fahmy, M. H., Abdou, M. A., and Darwish, M. A. (1999). Integral equations and potential-theoretic type integrals of orthogonal polynomials. *Journal of Computational and Applied Mathematics*, 106:245–254.

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of Integrals, Series, and Products.* Academic Press, New York-London.

Grenander, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.*, 1:195–277.

Hajek, J. (1956). Linear estimation of the mean value of a stationary random process with convex correlation function. *Czechoslovak Mathematical Journal*, 6(81):94–117.

Harman, R. and Štulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika*, 72(2):281–294.

Kanwal, R. (1997). *Linear Integral Equations.* Birkhauser, Boston.

Kiefer, J. (1974). General equivalence theory for optimum designs (Approximate Theory). *Annals of Statistics*, 2:849–879.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremem problems. *Canadian Journal of Mathematics*, 12:363–366.

Kiselak, J. and Stehlik, M. (2008). Equidistant and d-optimal designs for parameters of ornsteinuhlenbeck process. *Statistics & Probability Letters*, 78(12):1388–1396.

Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Annals of Mathematical Statistics*, 39:70–75.

Mason, J. C. and Handscomb, D. C. (2002). *Chebyshev Polynomials.* Oxford University Computing Laboratory, England, UK.

Müller, W. G. and Pázman, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika*, 90:423–434.

Näther, W. (1985a). *Effective Observation of Random Fields.* Teubner Verlagsgesellschaft, Leipzig.

Näther, W. (1985b). Exact design for regression models with correlated errors. *Statistics*, 16:479–484.

Pázman, A. and Müller, W. G. (2001). Optimal design of experiments subject to correlated errors. *Statistics and Probability Letters*, 52:29–34.

Pukelsheim, F. (2006). *Optimal Design of Experiments.* SIAM, Philadelphia.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Sympos., Univ. California*

Press, Berkeley, Calif., pages 355–372.

Sacks, J. and Ylvisaker, N. D. (1966). Designs for regression problems with correlated errors. *Annals of Mathematical Statistics*, 37:66–89.

Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. *Annals of Mathematical Statistics*, 39:49–69.

Schmidt, K. and Zhigljavsky, A. (2009). A characterization of the arcsine distribution. *Statistics and Probability Letters*, 79:2451–2455.

Torsney, B. (1986). Moment inequalities via optimal design theory. *Linear Algebra Appl.*, 82:237–253.

Zhigljavsky, A., Dette, H., and Pepelyshev, A. (2010). A new approach to optimal design for linear models with correlated observations. *Journal of the American Statistical Association*, 105:1093–1103.

Fakultät für Mathematik
Ruhr-Universität Bochum
Bochum, 44780, Germany
E-mail: holger.dette@rub.de

Institute of Statistics
RWTH Aachen University
Aachen, 52056, Germany
E-mail: pepelyshev@stochastik.rwth-aachen.de

School of Mathematics
Cardiff University
Cardiff, CF24 4AG, UK
E-mail: ZhigljavskyAA@cf.ac.uk